DOCUMENT RESUME

ED 031 777 24 EA 002 407

By-Bormuth, John

Comments on Professor Alkin's Paper Entitled "Evaluating the Cost-Effectiveness of Instructional Programs."

California Univ., Los Angeles. Center for the Study of Evaluation of Instructional Programs.

Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.

Report No-CSE-R-27

Bureau No-BR-6-1646

Pub Date May 69

Contract-OEC-4-6-061646-1909

Note-8p.; From the Proceedings of the Symposium on Problems in the Evaluation of Instruction (Los Angeles, December 1967).

EDRS Price MF-$0.25 HC-$0.50

Descriptors-Academic Achievement, *Cost Effectiveness, Decision Making, *Evaluation Techniques, Financial Support, *Instructional Programs, *Public Policy, *Test Construction, Test Validity

Evaluation of the ratio of cost to benefit of instruction must play an important part in the formation of a public policy on education. However, it is doubtful if evaluation is sufficiently developed to play such a role, because evaluation is based on student responses to test items which contain an indefinite bias and cannot be accepted as reliable data. A more reliable theory of test writing must be developed before evaluation can perform an important role in the making of public policy. Related documents are EA 002 475 and EA 002 534. (MLF)

ED031777

BR-6-1646
PA-24
OE-BR

COMMENTS ON PROFESSOR ALKIN'S PAPER ENTITLED "EVALUATING THE COST-EFFECTIVENESS OF INSTRUCTIONAL PROGRAMS"

John Bormuth

University of Chicago

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSE Report No. 27, May, 1969
University of California, Los Angeles

EA 002 407

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

CO-DIRECTORS

Merlin C. Wittrock Erick L. Lindman

ASSOCIATE DIRECTORS

Marvin C. Alkin Frank Massey, Jr. C. Robert Pace

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds.

COMMENTS ON PROFESSOR ALKIN'S PAPER ENTITLED "EVALUATING THE COST-EFFECTIVENESS OF INSTRUCTIONAL PROGRAMS"

John Bormuth

I think Dr. Alkin takes a very interesting, if not actually daring, position before this group when in his paper asserts that the evaluation of instruction is not really complete unless it includes some assessment of the costs involved in instruction and is not simply a measure of the behavioral outcomes of that instruction. In other words, to serve sufficiently the purpose of making educational decisions and educational policies, evaluation must also provide us with a description which expresses the benefits of instruction in relation to its costs.

My first reaction to this proposition is to give my hearty endorsement to the general idea, for the reason that to reject it would be approximately equivalent to saying that the values and products achieved through education are somehow exempt from competing with all our other values for their fair share of the public resources. We must realize that education must compete with space projects, war, cosmetics, and dog food for the citizen's dollar.

If the taxpayer has in the past been unwilling to provide education with all the financial support that some of us thought was its due, perhaps his hesitancy is attributable to the natural

aversion to buying a pig in a poke. Asking the taxpayer to put out money year after year in ever-increasing amounts for vaguely described products having even more vaguely described costs attached to them places a rather heavy strain on his credibility. Perhaps we should marvel at the fact that the taxpayer is so generous, rather than complain at his seeming parsimony.

If we really believe that public policy should be based upon informed public opinion, if we really believe that people have the right to know the effects that any treatment has upon them, and if we really believe that people have a right to know how their money is being spent, then we must agree with Dr. Alkin that the cost-to-benefit ratio of instruction must somehow be assessed. Hence, Dr. Alkin is asking us to reject the narrower conception of the role of evaluation. He is advocating that evaluation can and must play an important part in the formation of a public policy on education.

I have nothing but applause for Dr. Alkin's contention that evaluation should play a role in the formation of public policy on education. However, I have grave doubts that evaluation is sufficiently developed to play such a role. My argument is that evaluation is based upon the observation of student responses to some sorts of tasks, which I will henceforth call test items, regardless of what their specific form might happen to be; these items are

derived by some obscure procedure, and they are then selected for inclusion in a test on the basis of authoritarian judgments of some sort. Hence, a test score necessarily represents what the test-writers and judges choose to measure and cannot be interpreted as a measure of what the instruction actually taught. I further claim that until we are able to specify objectively the population of all possible test items that can be constructed for a given course of instruction, and that until we have developed a set of rules for mechanically deriving these questions directly from the instructional stimuli themselves, evaluation cannot provide us with information which can have sufficient scientific validity to meet the requirements implied by Dr. Alkin's proposal.

No one familiar with test-making procedures would seriously challenge the statement that the items which go into a test are selected solely on the basis of authoritative judgment of their relevance and their importance to the instruction. What seems to have been overlooked frequently is the fact that this method of test item selection makes the information from such tests unacceptable for the serious purpose of making public policy because its result is that the test scores tell us only what the test-makers want us to know. We have no way of determining what all the other things were that could have been taught by the instruction, nor can we even be certain that any performance on the items actually stemmed from the instruction presumably being examined. Therefore, we

must regard test items as containing an indeterminate bias. As a result, we can not accept test results as reliable data upon which to base decisions of public policy.

It seems necessary for us to conclude, then, that evaluation techniques can never perform an important role in the making of public policy. Now, is this so? I think not; but before they can do so, new techniques must be developed. These techniques must have the following characteristics:

1. They must permit us to enumerate exhaustively the behaviors that can be acquired as a result of exposure to a course of instruction. The resulting knowledge will allow us to inspect thoroughly the effects of a program and to draw a set of items that will enable us to examine an unbiased sample of those behaviors.

2. Our test construction techniques must permit us to derive the test items in a mechanical and completely reproducible process.

3. If taxonomic classifications are to be used in any way to describe the items so derived, these taxonomic classes must be defined in terms of the transformations by which they were derived from the instructional stimuli.

Although the meeting of these requirements may sound like an impossible task, it seems that the goals can be realized in a fairly adequate way. I suggested one possible solution in a paper I did for the Research and Development Center at UCLA a year ago. I

might add that Professor Anderson's transformations the other day seemed to be hitting very much at the same sort of solution.

I began my work with the statement that the knowledge transmitted by a course of instruction may be regarded as a closed system of statements phrased either in natural language or in some other symbolic system governed by syntactic constraints. When the instruction can be cast into this form, many of the test questions which are ordinarily constructed are expressible as transformations of the sentences occurring in the instruction. For example, suppose in the instruction we have a sentence of this sort: "High mountains tend to exhibit rapid hydraulic erosion." From this statement, by a specifiable transformation, we can produce the question, "What kind of mountains tend to exhibit rapid erosion?" Or still another question, by a slightly different set of transformations, would be, "What kind of geologic feature tends to be affected by the destructive forces of runoff?" There are a number of other questions that could be derived from exactly the same sentence, each of which would be derived by a slightly different transformation. These are innumerable in every sense of the word, and they are objectively derived.

The questions derived in this manner are not just those ordinarily judged to be measuring acquisition of explicitly stated facts but also include questions measuring various degrees of generalization and transfer. Further, these classes of questions are

objectively definable within the system of transformations used, but in general the questions so derived by the particular set of transformations I have been talking about deal with what we ordinarily classify as explicitly stated facts.

More recently, however, I have begun analyzing the syntactic constraints existing between sentences. These analyses seem to be leading to an ability to deal with questions commonly judged to be testing "knowledge of higher level concepts and more complex processes." Indeed, I seem to be getting the intuitively satisfying result that the traditional essay question has a generic kinship to the mundane short answer completion question. The two types of questions simply represent transformations operating at different levels in the syntactic structure of the discourse.

Many of Anderson's questions appear to fall within the classes derived by these transformations. But some of them also appear to represent transformations of an order that differs from any I had yet thought about.

What I am arguing, then, is that we need a theory of test writing and that until we have such a theory, the practical use of evaluation for the formation of public policy does not seem to me to be possible.